

## **Identification of the mode of evolution in incomplete carbonate sequences**

By Hohmann, Koelewijn, Burgess, and Jarochowska

This manuscript uses forward computer models to assess the fidelity of evolutionary signals preserved in stratigraphic successions forming on carbonate platforms. The motivation for the manuscript is the tendency of evolutionary (paleo)biologists to interpret the evolutionary modes of lineages at face value, rather than considering how the incompleteness of rock successions affects the preservation of evolutionary patterns. The authors examine trait lineages evolving under several evolutionary modes in two model scenarios using a simulated sea-level (SL) curve and the Pleistocene SL curve. An important finding is that the correct mode of evolution is not recovered from the modeled scenarios, and there is low statistical support for model recovery in general.

Carbonate platforms are important sources of the fossil record but have, to my knowledge, received less attention than passive margins from workers using models to analyze the preservation of stratigraphic and fossil records. I think this manuscript is compelling and well designed, and I have a few comments and suggestions. My expertise is not in evolutionary models, so my comments pertain mostly to the stratigraphic implications of the models.

### **General comments:**

Overall, the manuscript is concise and reads well, although the authors should proofread the manuscript thoroughly to check for minor grammatical issues such as subject–verb agreement and to check the formatting of citations and references.

The authors begin the manuscript with a discussion of stratigraphic bias on the fossil record. I suggest finding an alternative phrase to refer to the control of stratigraphy on the fossil record, as “bias” has negative connotations that may serve to justify the perceived shortcomings of paleontological investigations. Revealing the structure of the record (Holland, 2017) is the key to convincing neontologists to pay attention to stratigraphy. This is an important goal of studies like this, and this manuscript will be a stronger contribution toward this goal if the authors choose an alternative phrasing.

Model design: I take it that an assumption of the model is that all lineages are assumed to have an equal chance of being sampled across facies (or environments, as they are represented in the model). The authors should clarify this point in the Methods, since facies preference is another factor contributing to the “incompleteness” of the fossil record. Perhaps a follow-up study would be to examine the fidelity of evolutionary models reconstructed for lineages with distinct facies preferences in carbonate platforms.

The model constructs lineage patterns based on sampling one synthetic column at a time. I am curious whether the reconstruction of the evolutionary modes improves from tracking lineages from composite columns, similar to how graphical correlation integrates stratigraphic or biostratigraphic data from multiple locations.

The majority of the Results and Discussion focuses on general trends from scenarios A and B, and the figures almost exclusively show output from scenario A. I think it is great that the authors used the actual SL record in scenario B, but it barely comes up in the text. It is then odd when output from scenario B appears in Figure 7. I suppose the authors feel it is a moot point because the detection of evolutionary modes was not ultimately affected by SL or stratigraphic completeness, but I would like to see more discussion of the different (or similar) implications of both scenarios. I also think there is a missed opportunity to not only emphasize the relevance of the model findings to the real-world record, but also to say something about reconstructions of trait evolution from the real record.

Because the SL curves in each scenario impart different frequencies and durations of hiatuses, I think it is worth adding more emphasis of the importance of this to the real-world record. When the importance of hiatus frequency and duration is discussed in the Results (section beginning on line 485), the differences between scenarios are hinted at but not explicitly stated. If the real-world SL curve imparts many short hiatuses (and a more continuous age-depth model), then the potential to measure real modes of trait evolution is perhaps not as bad as we tend to fear because the record is “incomplete.” This point is eventually made later (lines 711–713). I think it is worth pointing this out to the reader earlier and briefly expanding on the implication that, although the real-world SL record is highly variable, it may result in a stratigraphic record that has a fairly good ability to preserve evolutionary change in the fossil record.

### **Tables and figures**

The figures are appropriate and support the text, although I suggest a few minor modifications to some of the figures. Additionally, some of the figure captions need more explanation of the information conveyed in the figures.

In Figure 2, it would be helpful to have a color key for the different facies depicted in the simulated shelves. Differentiating among facies is not a focus of this manuscript, however, they are clearly shown in this figure and in Figure 3. In the caption, please clarify what are “the graphs,” or refer to other figures by number.

In Figure 3, please include a color key to the facies, as suggested for Figure 2. In the caption, I suggest either stating the explicit distance from shore or adding a line to the Wheeler diagrams to show where the graphs were extracted, rather than referencing the “middle” of the grid. Please clarify: the extracted “graphs” are 2C and 2F?

Figure 7 illustrates the point that stratigraphic completeness in and of itself has little bearing on the preservation of trait evolution patterns. It is unclear why a column from 2 km in scenario A is compared to a column from 6 km scenario B. Scenario B is so little discussed in the Results and Discussion, and there is little explanation for the rationale of comparing two different platform locations across the two scenarios. I think the point is that completeness doesn’t vary too much, but it is hard to take that away from comparing two completely different column locations and scenarios. Please explain this choice in the caption and text or consider modifying the figure to show outputs from more readily comparable example columns.

Figures 8 and 9, 10: Captions need explanation of the abbreviated tested modes in the legends.

I suggest adding a table to report the results of the AIC analyses for the different simulations and models. The pertinent results are reported in the text, but a table would be helpful for presenting all the AIC values.

### **Referred references**

Holland, S.M., 2017, Presidential address: Structure, not bias: *Journal of Paleontology*, v. 91, p. 1315-17.

Holland, S.M., 2022, The structure of the nonmarine fossil record: Predictions from a coupled stratigraphic–paleoecological model of a coastal basin: *Paleobiology*, p. 1-25, doi: 10.1017/pab.2022.5

### **Specific comments**

Line 19: The clause, “computer simulations of geological processes . . .” needs an object for “allow.”

Line 29 and elsewhere: What is meant by “adequate model”? The authors mention “adequate models” several times in the text before getting close to what they mean on pages 32 and 36. Is an adequate model highly supported statistically? Does an adequate model faithfully simulate the mode of evolution? Please define this term in the Introduction and add a brief explanation in the Abstract.

Line 71: This phrasing makes it sound like sedimentology and stratigraphy are the only disciplines that are jargon laden. It is a fair point that sed/strat—and many disciplines—are often written for narrow audiences that may make it difficult for outsiders to glean relevant information. I think the issue is more that evolutionary biologists are typically interested in *ages* of fossils that can be used to calibrate trees, not their stratigraphic contexts. They don’t know that the stratigraphy informs the age estimate and they do not know to look for it.

Line 94: Fossil-bearing stratigraphic successions.

Line 203: Were run

Line 289: Awkward wording

Line 295: “. . . trait value that describes”

Lines 658–660: Yes, it is important to examine multiple columns along dip for interpreting the patterns. The authors may want to acknowledge here that this is a bit of an oversimplification that may not directly apply to carbonate platforms. Given that the geometry of carbonate platforms causes hiatuses to form along dip, similar age-depth models should be recovered from neighboring columns. In practice, it may be difficult to compare platform locations to the distal slope locations, and I expect the facies and fossil assemblages would be very different. During lowstands, sediment may accumulate on the slope, but it will be redeposited material that may not preserve fossils well, making it difficult to sample the traits of interest.

Line 691: “. . . jumps will coincide with gaps . . .”

Line 716: “First” (avoid adding -ly in writing)

Line 723: “Second”

Line 778: I’m confused by the use of “ground truthing” here or by the wording of this sentence. Are the authors suggesting that the models offer ground truth? Ground truth can only be gotten in the field.

Lines 783–784: see also strataR in Holland (2022)